

# An Intelligent Bay Geo-information Retrieval Approach based on Geo-ontology

Rulin Xiao, PhD student

Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences  
Beijing China, 100101  
xiaorl@lreis.ac.cn

Yunyan Du, Associate professor

Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences  
Beijing China, 100101  
duyy@lreis.ac.cn

Wei Liu, PhD student

School of Environment Science and Spatial Informatics, China University of Mining and Technology  
Xuzhou China, 221116  
liuwei@lreis.ac.cn

Yawen He, PhD student

Yantai Institute of Coastal Zone Research for Sustainable Development, Chinese Academy of Science, Yantai China, 264003  
heyw@lreis.ac.cn

**Abstract.** In the era of information explosion, information retrieval has become a bottleneck in information sharing and integration. However currently, the existing information retrieval methods are mainly based on keyword matching, which can not fully take advantage of the information context and potential knowledge. All of these methods are particularly inefficient as to geospatial information which is more complex and unstructured. Nevertheless, geospatial ontology (Geo-ontology) has been used to enrich geospatial objects with semantic information which could be very helpful in the geospatial information retrieval and integration. And there is a wealth of relations between objects (such as the relation between bay and the inter-tidal zone) in bay field. Thus in this paper, an intelligent bay geo-information retrieval approach based on bay geo-information ontology has been proposed. Firstly, the procedure of establishing the bay geo-information ontology database is introduced. Secondly, both the intelligent retrieval mechanism and approach in bay geo-information are expatiated, which are used to convert the user retrievals request to semantic request and map the semantic request to concrete query of actual data source by adopting the knowledge that 's explicitly expressed in formal ontology modeling language (OWL). Finally, the method was applied in the bay information management system, and the feasibility and efficacy of the geo-information retrieval approach is well verified.

**Keywords:** *Geo-Ontology; Information Retrieval; Bay Geo-Information System*

## I. INTRODUCTION

In the era of information explosion, information retrieval is the key point in the process of geo-information sharing and integration, and always become a bottleneck. So Open GIS Consortium (OGC) defined the geo-information retrieval framework and standards in its interoperability specification [1], and recommended Metadata Catalog Service as the mainstream geo-information retrieval method. However, Metadata Catalog Service is mainly based on keyword matching, which is limited to predesigned pattern, only several key words are allowed for the user to represent a search, and

just does keyword or sub-string matching in its internal searching process. So its result could be some general information. As it can neither take full advantage of information context, nor can it make use of the deep and hidden knowledge, the result could be null or be wrong, which can not meet the users' need [2]. Especially in case of geo-information field, because of complex object and unstructured data, it can not use simple key word to search, and a novel approach is required.

Ontology can enrich semantics, which is useful in the information retrieval, integration and mining and can help to improve the recall rate and its accuracy. In marine field, there is rich in semantic characters and relations. Take objects in bay field for an example: if the user wants to retrieval information of Bay A, but there is only metadata for the data of land area, sea area and intertidal zone of it in the geo-data exchange center, and each is composed by several subset data. If key word matched approach was used, the user inputs "Bay A" and can not get any result. But actually, bay was composed by land area, sea area and inter-tidal zone, they all have already existed in the database, and should be returned together with the relation between them and the ones that the user wanted explicitly represented. In contrast, another case is that, if the user wanted to get the data of Bay A, while the geo-data exchange center has all data of bays lie in South China Sea and with metadata as a whole. If key world based search, he can not get what he needed, neither. But actually, it can be subtracted from the data already exist. Why should it be this? Because the key word based matching can not know the relation between the data searched for and the data supplied. So ontology is necessary.

In September 2003, the State Council formally approved the "Comprehensive Survey and Evaluation of China Offshore Marine" special project (referred as the "special project 908"). According to overall strategic concept of China's "digital Ocean" concept and objectives of the project 908, marine geo-information sharing is important and critical. As the forefront

of marine, bay has superior geographical location and unique natural environment, play a key role in the national economy and people's livelihood [3]. So in this paper, an intelligent geo-information retrieval approach based on geo-ontology was proposed to improve the geo-information retrieval rate and its accuracy. And this research mainly focused on the bay information retrieval.

In geo-information field, there have been many researches and applications on ontology and have gain some achievement [4-12]. However, those researches and applications often only simply introduced the concept of ontology, hardly have considered combining it with the unique characteristics of geo-information. They are all limited to the use of relation and knowledge between non-spatial properties of geo-information; few of them have take advantage of spatial relation and knowledge. Related work in marine field is even less. Findler and Malyankar established the corresponding ontology for coastal objects in their research [13]. Wariyapola and Patrikalakis researched in the standard and technology in the metadata and ontology creation process [14]. Yuyan Du and Dandan Zhang proposed a good approach to organize bay geo-

data based on geo-ontology [15], but how that kind of ontological databases should be used and what application could it be used in have not been get further research in their research. Jinggui Wang and Fenzhen Su introduced a kind of geo-knowledge query approach based on ontology and its application in coast zone resource and environment system [16], but that approach only used the simple relations between concept, property and instance, without using the complex relations such as equivalent, disjoint, parent-son, containing and being contained relation.

## II. REPRESENTATION AND ESTABLISHMENT OF BAY GEO-ONTOLOGY

Firstly, this research introduced an ontological thought and constructed the bay geo-ontology concept model which can fully reflect the hierarchical structure of geo-objects; Secondly, under the guidance of the bay concept model, the ontological geo-database was established based on the mapping between bay concept and bay geo-data; finally the integrated management of sea, land and beach was realized, which made the bay geo-information sharable and interoperable.

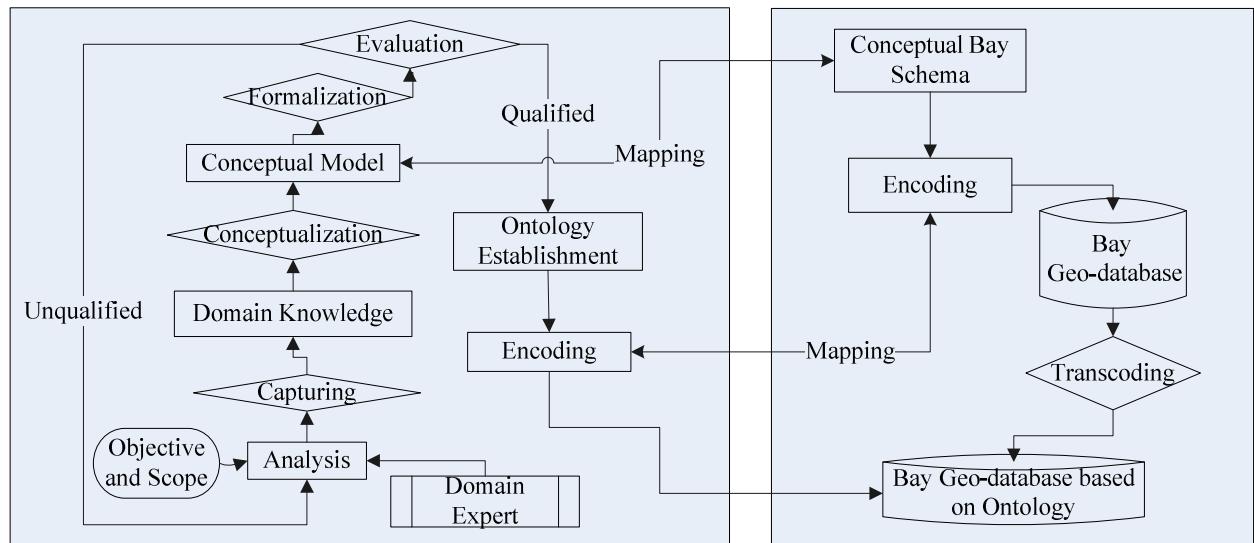


Figure 1. Framework for building bay ontology

### A. Representation of Bay Geo-ontology

Geo-ontology is defined as common shared conceptualized model in specific geo-information community or explicit formalized specification in the domain perception world [17], while bay geo-ontology is defined as the ontology related with the research domain of bay. After review of advantages and disadvantages of ontology representation models such as triple model, quintuple model and sextuple model [18-20], this research proposed quadruple model as the representation model for bay geo-ontology. And OWL (Web Ontology Language) is chosen as the model language. It details as follows:

- Domain Ontology: It is defined as a quadruple  $O = (CD, RD, AD, rD)$ . In it,  $O$  is domain ontology set,  $CD$  is a concept set in the specific domain,  $RD$  is the set of binary relation between concepts such as property relationship between two concepts;  $AD$  is the property

set;  $rD$  is the set of semantic relation between concepts which is used to describe relations between concepts. These semantic relations include equivalent relation (same-as), inheritance relation (is-a), antagonism relation (inverse), transitivity relation (transitive), symmetry relation (symmetry), whole-part relation (part-of), intersection relation (intersect) and spatial relations such as topological, directional and distance relation.

- Domain Concept: It is defined based on domain ontology as  $DO = \{(D_1, DD_1), (D_2, DD_2), \dots, (D_n, DD_n)\}$ . In it,  $DO$  is domain concept set,  $DD_i$  is domain description,  $D_i$  is domain key word.
- Global Ontology: It is defined based on domain ontology as a quadruple  $SO = (C, R, A, r)$ . In it,  $C$  is a concept set in domain,  $R$  is the two-way binary relation

between concepts, A is the property set, r is the set of semantic relation between concepts which is used to describe relations between concepts and its semantic rule is consistent with domain ontology's.

### B. Establishment of Bay Geo-ontology

The establishment of bay geo-ontology is a process in which according to the actual purpose and application requirement, each concepts, their relations and properties should be explicitly formally described and encoded by the way introduced in the Part A.

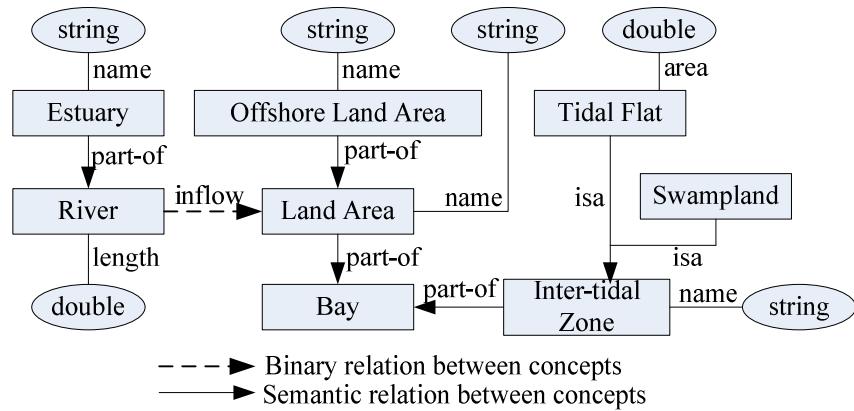


Figure 2. Domain ontology of bay( only partial)

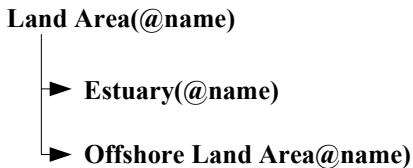


Figure 3. XML data source (source1)

**T1:**Land Area->source1/Land Area:v1  
**T2:**name->v1/@name:v2  
**T3:**inflowed\_by,River->v1/River:v3  
**T4:**composed\_by,Estuary->v3/Estuary:v4  
**T5:**Estuary\_name->v4/name:v5  
**T6:**composed\_by,Offshore Land Area->v1/Offshore Land Area:v6  
**T7:**Offshore Land Area\_name->v6/name:v7

Figure 4. Mapping of source1 from global ontology to local shcema

- Ontology analysis according to the actual purpose and application scope. By collection and collation of bay-related materials such as literature, theories, research reports and atlases, and combining domain experts' knowledge, the bay-related glossary was gotten. The glossary contains term such as bay, tidal zone, offshore land area, bay mouth and coastline, and defines the meaning of each term and relation between them. In this paper, lots of experts in marine field had taken part in the analysis process.
- Identifying the relation between concepts in the glossary, and defining domain knowledge and representing it in natural language. Term that can be defined by intention should be expressed. Intention definition is referred as using only a limited number of inextricably properties of it to define the term itself.
- Formalizing those definitions in formal ontology language. Firstly, all concept related with domain knowledge were defined and the ontology concept set was formed, then based on the relations between

The establishment of geo-ontology model is complex and time-consuming process. It needs lots of domain experts to participate in this work. In this paper, according to software engineering philosophy, based on combining advantages of skeleton model and methontology model [21-24], we took a spiral model to build ontology. The whole process contains 5 phases: analysis, representation, evaluation, establishment, coding. The overall framework is as shown in Fig. 1. The box in the left depicts the establishment process, and the one in the right depicts the geo-data organization process. It details were introduced as follows and the bay geo-ontology had been established is as shown in Fig. 2(only a small part).

ontology concept and spatial relations between object, these ontology concept were organized well and the classification system of the ontology was built up.

- Evaluating the correctness and effectiveness of each ontology. The evaluation principles are primary clarity, objectivity, consistency, integrity, scalability, and minimal commitment [25]. If ontology is not qualified, it should be redone from the beginning.
- Coding the ontology. Because quadruple model is used as the ontology representation model, it can not model instance which is corresponding to the data record in database. Thus how could be these ontologies linked with those data records in database? A joint ID should be used though which the semantic information and the data can be linked together (And the relationship between the ontology and the data records in database is 1 : n). So ontology needs to be encoded with unique ID code. The code will be used in the mapping between the ontology and the data.

### III. INTELLIGENT RETRIEVAL MECHANISM BASED ON GEO-ONTOLOGY

In face of a large number of heterogeneous, user customized bay data resource in distributed environment, without a good mechanism it is hard to reach dynamic discovery, retrieval and integration. Usually, retrieval may involve more than one model, if tradition retrieval approach is used, it needs to large-scale search over multi-nodes, and its inefficiency is obvious and can not distinguish information of different domain. So in this paper, firstly based on the thought of managing resources in different domain with different method, the domain ontology was used to provide consistent semantic specification for data source in the same domain and was used to define domain concept and global schema; secondly the mapping from heterogeneous data to global schema was used to transform data using XML; finally the intelligent retrieval approach based on ontology was realized. This method can facilitate the management of data resource of different fields, and can reduce the total number of hops that needs to search through to alleviate the system load and also improve the retrieval efficiency and recall rate. As shown in Fig. 2, it is the schematic view of ontology in bay (only partially). When retrieval in bay field, it can be used as the global schema. By reasoning based on the semantic rules and knowledge, it realizes such kind of ontology based intelligent retrieval approach.

#### A. XML Based Mapping from Heterogeneous Data to Global Schema

Firstly, the retrieval request from user interface is sent to the global retrieval engine; secondly the request was converted to retrieval based on global ontology, finally by mapping between ontology and xml, the request was converted to xml based retrieval. The mapping between Ontology and XML is primary on rules as follows:

- Complex XML element is mapped to ontology concept, corresponding to the non-leaf node in the ontological concept diagram;
- The leaf node in the ontological concept diagram is property of concept, corresponding to the simple type data in XML file. The dual relation between concepts is used to represent relation between complex type data in XML file.
- By the concept abstraction using global ontology to achieve mapping between ontology and xml file, which hides the heterogeneity between different data source and provide a unified retrieval interface.

The mapping rule is defined as  $T:p \rightarrow r/q:v$ . In it,  $T$  is tag for rule,  $r$  is rule's root, and  $q$  is the local (relative) path of XPath,  $p$  is the path of ontological schema. The rule's root( $r$ ) can not only be variable and also be URI. If it was a variable,  $p$  is role path; if not,  $p$  is concept path. If it was a variable,  $t$  is referred as relative mapping rule; if not,  $t$  is referred as absolute mapping rule. For example, the data source source1 (As shown in Fig. 3), is mapped from global ontology (As shown in Fig. 2) to local schema is (As shown in Fig. 4).

#### B. Ontology Based Intelligent Retrieval Approach

The rich semantic information provided by bay ontology and reasoning ability enhances the user's retrieval with powerful intelligence. As shown in Fig. 2, as part of land area, off-shore land area has all properties of land area. Tidal flat and swampland are subclasses of tidal zone, retrieval of tidal zone can automatically be extended to all its subclasses.

The ontology based intelligent retrieval process is as follows:

- After receiving retrieval request based on concept (global schema) from user, the concept set  $C = \{c_1, c_2, \dots\}$  was generated, which consisted of all related concept according to semantic relation between ontological concept.
- According its semantic relation, all concepts in set  $C$  should be standardized and generalized. By standardization of Equivalent concept and expansion of relations such as parent-child relation, containing relation and being contained relation to expand the user's retrieval request.
- Establishing relation set  $R = \{r_1, r_2, \dots\}$  of dual relation between the concepts in set  $C$ .
- Inferring based on all relations in relation set  $R$  according rules predefined such as using inverse, symmetry and transitivity transformation and reasoning to get more deep semantic relation.
- Through these four pre-processing steps, it generates a new semantic extended retrieval;
- Decomposing retrieval into multiple sub-retrieval set and to discovery data records based sub-retrieval pattern.
- Sum up all sub-retrieval results as the final output.

It uses mapping between ontology and the XML schema of the data source, and then rewrites the retrieval based on global schema to retrieval based on XPath of different data resource, finally generate the retrieval execution schedule.

### IV. USE CASE

In this section, we applied our bay geo-data organization framework and retrieval mechanism in the bay geo-information system for Perl River Delta. We took recall rate and accuracy as the evaluation index. Their definitions are as follows:

$$R = Nx / N' \quad (1)$$

$$A = Nx / N \quad (2)$$

$R$  is the recall rate,  $A$  is the accuracy.  $Nx$  is the number of correct data records,  $N'$  is the number of related data records sand  $N$  is the number of data records have been returned by the retrieval.

Through test in the system and by rough statistics (As different retrieval input will have different retrieval result), we found that compared to traditional retrieval mode, the

intelligent retrieval mode based on ontology have obviously improved both the recall rate and the accuracy. And the improvement in the accuracy was not remarkable in the recall rate. Because besides traditional retrieval results, it returned the ones that have semantic relation with that user request more or less.

## V. CONCLUSIONS AND FUTURE WORK

In this work, a bay geo-data organization method based on ontology has been described, and it works well in building up a semantic geo-data base. On that basis, it proposed in a kind of intelligent retrieval mechanism. In the retrieval process, it uses domain ontology as global schema and uses ontological knowledge and logic rule-based reasoning to overcome the semantic heterogeneity. The retrieval request was first semantic extended and is rewritten when the actual data source's schema is not same as the global schema. And sometimes, retrieval decomposition is also necessary. Finally, both the geo-data organization method and the intelligent retrieval mechanism had been applied in real application, and their feasibility and effectiveness had been verified. Its advantages are as follows:

- Based on human being's cognitive model, using object-oriented idea makes the data more natural and better organized.
- It enables integration and interoperability of different data source. The global ontology and mapping mechanism play a key role in it.
- By that intelligent mechanism, it not only hides the complex details of data source from the user, but also makes the retrieval operation simpler and more convenient.
- It can improve the accuracy and recall rate of retrieval.

As a future work, we are working on three aspects.

- Firstly, on the aspect of distributed data source. Our present work is on basis of data in local database and is not suitable for data in the distributed network environment. Service oriented architecture (SOA) and semantic web are the mainstream development. How to combine that mechanism with those technologies and philosophy to work well in the distributed environment? There needs further work.
- Secondly, on the aspect of deeper semantic level. Present work uses quadruple model to represent ontology. Instance ontology have not been take into consider and thus knowledge between instance can not been used. It is a semantic model on metadata level or data schema level. The next step will focus on semantic enriching in data level, which would be more laborious and time-consuming.
- Finally, on the aspect of making the geo-ontology more geospatial enabled. Our present work only simply converts spatial relations or knowledge to non-spatial property. For example the intersection relation between River and Bay is simply converted to ontological property - "inflow", that is not enough for a user who

want to know where is it flew into the bay, in the west or in the east? So an advanced representation model is needed to express geospatial knowledge and relations (such as distance relation, directional relation and topological relation) well. On that basis, spatial reasoning and mining can be possible and feasible.

## REFERENCES

- [1] OGC. The OpenGIS Abstract Specification. Topic 13: Catalog Services (version 4), 1999.
- [2] Egenhofer M J. Towards Semantic Geospatial Web. GIS'02, McLean Virginia USA, 2002.
- [3] Sangyun Wu and Baoyin Liu. Fundamentals of China's bay Information system. Ocean press, 2007.
- [4] Hardy Pundt and Yaser Bishr. Domain ontologies for data sharing—an example from environmental monitoring using field GIS. Computers & Geosciences, 2002, 28: 95–102.
- [5] Harry T. Uitermark, Peter J.M. van Oosterom, Nicolaas J.I. Mars and Martien Molenaar. Ontology-based integration of topographic data sets. International Journal of Applied Earth Observation and Geoinformation, 2005, 7:97–106
- [6] Frederico Fonseca, Max Egenhofer, Clodoveu Davis and Gilberto Câmara. Semantic granularity in ontology-driven geographic information systems. Annals of Mathematics and Artificial Intelligence, 2002, 36: 121–151.
- [7] Alia I. Abdelmotti, Philip D. Smart, Christopher B. Jones, Gaihua Fu, and David Finch. A critical evaluation of ontology languages for geographic information retrieval on the Internet. Journal of Visual Languages and Computing, 2005, 16: 331–358.
- [8] Isabel F. Cruz, William Sunna, Nalin Makar, Sujan Bathala. A visual tool for ontology alignment to enable geospatial interoperability. Journal of Visual Languages and Computing, 2007, 18:230–254
- [9] David Mallenby and Brandon Bennett. Applying spatial reasoning to topographical data with a grounded geographical ontology. GeoS 2007, LNCS 4853:210–227.
- [10] Peng Yue, Liping Di, Wenli Yang, Genong Yu and Peisheng Zhao. Semantics-based automatic composition of geospatial Web service chains. Computers & Geosciences, 2007, 33: 649–665.
- [11] Xu Quan-Li, Yang Kun, Wang Jun, Peng Shuang-Yun and Yi Jun-Hua. A study on ontology-driven geospatial-information retrieval systems in the semantic web. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Beijing, 2008, Vol. XXXVII. Part B2: 733–738.
- [12] Peisheng Zhao, LipingDi, GenongYu et al. SemanticWeb-basedgeospatialknowledge transformation. Computers & Geosciences, 200, 35:798–808.
- [13] Findler,N.V and Malyankar,R.M. Digital goverment: an ontology for geospatial knowledge. NSF Award EIA 9876604, 1999.
- [14] P.C.H.Wariyapola, N.M.Patrikalakis, S.L.Abrams and P.Elisseeff. Ontology and Metadata Creation for the Poseidon Distributed Coastal Zone Management System. IEEE Forum on Research and Technology Advances in Digital Libraries, 1999, 180–189.
- [15] Wang Jinggui, Su Fenzhen, Zhou Chenghu, Du Yunyan and Yang Xiaomei. Oceanographic ontology-based spatial knowledge query. Acta Oceanologica Sinica, 2005, 24(4):66~71.
- [16] Du Yunyan, Zhang Dandan, Su Fenzhen, Yang Xiaomei and He Huizhong. Geospatial data organization of the bay based on the geo-ontology—a case study of the Liaodong Bay. Geo-information Science, 2008, 10(1): 7-13.
- [17] Wang JingGui. Geospatial data integration based on geo-ontologies. PhD thesis, 2005.
- [18] Perez A G, Benjamins V R. Overview of knowledge sharing and reuse components: ontologies and problem-solving methods. Proceedings of the IJCAI-99 workshop on ontologies and problem-solving methods (KRR5), Stockholm Sweden, 1999, 1-15.

- [19] Myo-Myo Naing, Ee-Peng Lim and Dion Goh Hoe-Lian. Ontology-based web annotation framework for hyperlink structures. Proceedings of the Third International Conference on Web Information Systems Engineering, 2002, 184-193.
- [20] Neches R, Fikes R E, Gruber T R et al. Enabling technology for knowledge sharing. *AI Magazine*, 1991, 12(3):37-55.
- [21] Uschold M and Gruninger M. Ontologies: principles, methods and applications. *Knowledge Engineering Review*, 1996, 11(2):93-136.
- [22] Uschold M and King M. Towards a methodology for building ontologies. IJCAI-95 Workshop on Basic Ontological Issues in Knowledge Sharing, Montreal, 1995, 41-64.
- [23] Fernández-López M. Overview of Methodologies for Building Ontologies. Proceedings of the IJCAI99 Workshop on Ontologies and Problem-Solving Methods:Lessons Learned and Future Trends, 1999, (4):1-4.
- [24] Fernández-López M, Gómez-Pérez A and Juristo N. METHONTOLOGY: from ontological art towards ontological engineering. AAAI-97 Spring Symposium on Ontological Engineering, Stanford, 1997, 33-40.
- [25] Gruber T R. Towards principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, 1995, 43:907-928.